M.A. Gedil
N.K. Steiner
C.K. Hurley

Genomic characterization of KIR2DL4 in families and unrelated individuals reveals extensive diversity in exon and intron sequences including a common frameshift variation occurring in several alleles

Key words:
cell-surface molecules; human; interspersed repeats; intron; LINE; molecular biology; NK cell; SINE

Acknowledgments:
This research is supported by funding from the Office of Naval Research N00014-00-1-0898. The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Navy, the Department of Defense, or the U.S. government. We thank Dr Phillip Posch for his invaluable comments and suggestions.

Abstract: The KIR2DL4 gene including a portion of exon 1 through exon 9 was sequenced from two families and eight cell lines from the International Histocompatibility Workshop (IHWS). Two known alleles and eight variants were detected. Overall, there were five synonymous and three non-synonymous changes when the variants were compared to the coding sequences of the most closely related known alleles plus a common frameshift change in five of the variant alleles. Alignment of the new variants with all known alleles showed that the regions encoding the extracellular region and the cytoplasmic tail were the most polymorphic. Two non-synonymous changes, P146H and L161V, occurred in an extracellular immunoglobulin-like domain. Five of the eight variants had a single adenine deletion in the exon encoding the transmembrane region, potentially resulting in a truncated protein lacking the cytoplasmic tail. The distribution of the deletion variant among many KIR2DL4 alleles may explain the high frequency of this variation in the population. Four of the eight consanguineous IHWS cell lines were found to be heterozygous for KIR2DL4 carrying two alleles that differed from one another by a few nucleotide substitutions. Analysis of intron sequences in the families revealed the nature and distribution of interspersed repeat elements which comprise 46% of the KIR2DL4 nucleotide sequence and consist of 12 elements including six SINEs (13.73% of the total length), one LINE (12.41%), and five LTR elements (19.51%). The results revealed the presence of extensive diversity in the KIR2DL4 gene. This is the first extensive report providing both exon and intron data in related individuals.

Authors' affiliation:

M.A. Gedil
N.K. Steiner
C.K. Hurley

Department of Oncology, CW Bill Young Marrow Donor Recruitment and Research Program, Georgetown University Medical Center, Washington, DC, USA

Correspondence to:
Carolyn Katovich Hurley, Ph.D.
E404 Research Building
Georgetown University Medical Center
3970 Reservoir Road, NW
Washington, DC 20057, USA
Tel.: +1 202 687 2157
Fax: +1 202 687 6440
e-mail: hurleyc@georgetown.edu

Human killer-cell immunoglobulin-like receptors (KIRs) play an important role in controlling natural killer (NK) cell function by delivering activating or inhibitory signals that depend, in part, on the detection of HLA ligands (1, 2). Encoded by a family of 15 genes, KIRs are characterized by the presence of two to three immunoglobulin (Ig)-like extracellular domains and from zero to two cytoplasmic immunoreceptor tyrosine-based inhibitory motifs (ITIM). In general, KIR with one or two ITIM and long cytoplasmic tails (KIR2DL1-5 and KIR3DL1-3) are inhibitory receptors, while KIR with short cytoplasmic tails containing no ITIMs, but with a charged

Received 11 November 2004, revised 7 January 2005, accepted for publication 18 January 2005

doi: 10.1111/j.1399-0039.2005.00380.x

Tissue Antigens 2005: **65:** 402–418

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2005	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Genomic characterization of KIR2DL4 in families and unrelated individuals reveals extensive diversity in exon and intron sequences including a common frameshift variation occurring in several alleles	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Department of Oncology, CW Bill Young Marrow Donor Recruitment and Research Program, Georgetown University Medical Center, Washington, DC**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) **Naval Medical Research Center 503 Robert Grant Avenue Silver Spring, MD 20910-7500**	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**SAR**	**17**	

residue (arginine or lysine) in their transmembrane regions (2DS1-5 and 3DS1) are stimulatory receptors. Although not all KIR ligands have been identified, some KIR inhibitory receptors detect the presence of specific HLA subtypes. Loss of these ligands by malignant or virally infected cells results in a loss of inhibition, resulting in susceptibility to NK cell cytotoxicity.

Among KIR, KIR2DL4 is unique in several features. At the DNA level, the KIR2DL4 gene consists of eight exons as exon 4 has been deleted (Fig. 1), while most other KIR2D genes consist of nine exons with a pseudoexon 3 (3, 4). Therefore, the extracellular Ig-like domains of KIR2DL4 are encoded by exons 3 and 5 (a D0–D2 configuration), respectively, unlike other KIR2Ds that possess a D1–D2 configuration (exons 4 and 5, respectively). The use of exon 3 rather than exon 4 is a feature shared only by KIR2DL5 (5). Unlike other KIRs that are expressed in a stochastic fashion in individual NK cells, the expressed KIR2DL4 alleles are apparently transcribed by every NK cell (6). Since the promoter regions of most KIR genes exhibit a high level of sequence similarity (>91% identity), it is thought that they may be regulated in a similar manner. However, the promoter region of KIR2DL4 shares only 69% identity with the same regions of other loci, and this divergence may account for KIR2DL4 expression in all NK cells (7).

The presence of a single inhibitory motif (ITIM) in KIR2DL4's cytoplasmic domain (other KIR2DLs possess two ITIMs) and the presence of a positively charged arginine in its transmembrane region (not present in other KIR2DLs) suggest both inhibitory (8) and activating functions (9, 10) for this unique KIR receptor. The phosphorylated single ITIM of KIR2DL4 recruits the Src-homology 2-containing protein tyrosine phosphatase-2 (SHP-2) leading to inhibition of NK cytotoxicity by dephosphorylating multiple targets in the immunoreceptor tyrosine-based activating receptor pathway (11). An early study suggested that KIR2DL4 binds the non-classical HLA-G molecule, which is expressed during pregnancy on fetal trophoblast cells and postulated that KIR2DL4 might play a role as an inhibitory receptor to block maternal NK cell-mediated attack of the fetus (12). However, a later investigation failed to detect any interaction between KIR2DL4 and HLA-G (13) leaving the identity of the KIR2DL4 ligand in doubt. As an activating receptor, KIR2DL4 may contribute to protection against infection via the induction of IFN-γ that is mediated by the transmembrane arginine (9, 10).

The KIR locus is very complex having arisen via multiple gene duplication and recombination events (7, 14, 15). Studies defining KIR gene content on chromosome 19 have elucidated over 35 haplotypes consisting of seven to 14 genes per haplotype (1). KIR2DL4 is one of the three conserved framework loci being present in all KIR haplotypes. Recent investigations using probe hybridization (16) and a quantitative polymerase chain reaction (PCR) technique (17) have shown that some haplotypes might carry multiple copies of the KIR2DL4 gene. To date, nine KIR2DL4 alleles have received allele designations (18). KIR2DL4 alleles have been described that contain a single nucleotide deletion at the end of the transmembrane domain (18–20). Further investigation of the function and expression of the putative truncated protein encoded by this KIR2DL4 mutation showed significantly reduced or lack of detectable membrane expression (10, 21). For other KIR genes, allelic variants altering the surface expression of the KIR molecule have also been observed (22–24).

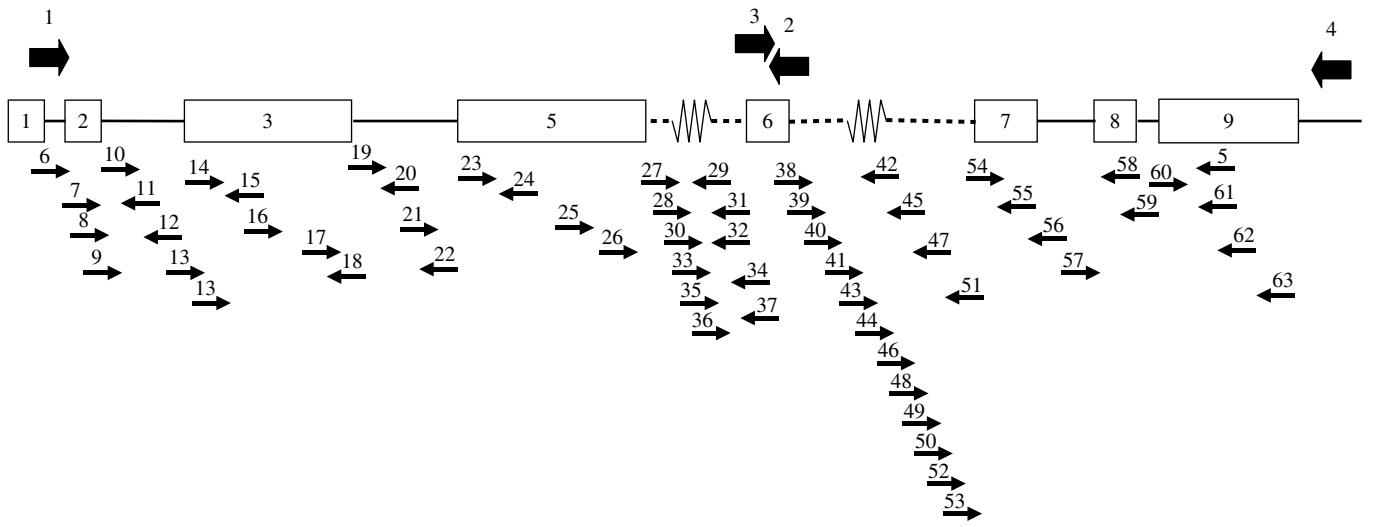

Fig. 1. **The approximate annealing positions of polymerase chain reaction and sequencing primers used for KIR2DL4 characterization.** Boxes indicate exons. The solid arrows on top indicate the positions of primers used to amplify the gene. The arrows below the exons show the positions of the sequencing primers. The numbers over each primer refer to the list of primers summarized in Table 1. The exons are numbered based on an alignment of all of the KIR gene sequences as reported in Garcia et al. (26).

To further investigate the complexity of the KIR locus, we characterized allelic polymorphism of the KIR2DL4 gene in two families (African-American and White) and eight unrelated individuals from the International Histocompatibility Workshop (IHWS) cell panel by sequencing a portion of exon 1 through exon 9 and all introns. In this study, we report the discovery of two known alleles and eight new variants. Interestingly, five of the eight new variants contain a single adenine deletion in the exon encoding the transmembrane region. Since the majority of KIR2DL4 sequences submitted to the GenBank are cDNA sequences, the goal of this study was to characterize KIR diversity and to gather intron information to facilitate KIR genotyping.

Materials and methods

DNA samples

Genomic DNA was isolated from Epstein–Barr virus-transformed B-cell lines derived from African-American family 004 (father GU2015, mother GU1183 and children GU1181, GU1182, GU1194, and GU1197) and White family 001 (father GU324, mother GU321, and children GU320, GU322, GU323, GU325, and GU326) using a QIAamp® DNA Blood Mini Kit (Qiagen, Valencia, CA) following manufacturer's instructions. Similarly, DNA was isolated from cell lines of eight consanguineous individuals from the International Histocompatibility Workshop (9005, 9010, 9023, 9030, 9035, 9042, 9050, and 9065 (25).

PCR amplification

The KIR2DL4 gene was amplified by PCR using two pairs of primers (2DL4E1F1 and 2DL4I6R1; 2DL4E6F2 and 2DL4I9R2; Table 1 and Fig. 1). Each pair was designed to amplify a part of the gene (exon 1 to intron 6 and exon 6 to exon 9, respectively) with an overlap of about 170 bp in the two amplicons. PCR primers were designed to include all known KIR2DL4 alleles (18). Hot Start PCR using Platinium® Taq DNA polymerase High Fidelity (Invitrogen Life Technologies, Carlsbad, CA) was performed to prevent non-specific priming and to increase the accuracy of copying this large gene segment. Each reaction (100 μl) consisted of 200–500 ng template DNA, 1× high-fidelity buffer (Invitrogen Life Technologies), 2 mM $MgSO_4$, 200 μM dNTP, 0.2 μM each of the forward and of the reverse primer pair and 2% DMSO. A modified reaction volume of 50 μl with 0.4 μM each of the forward and reverse primer pair and no DMSO was used for the PTC-225 thermal cycler. Amplification reactions were performed in a model 9600 (Perkin-Elmer, Norwalk, CT) or a PTC-225 (MJ Research) thermal cycler. The amplification conditions for the model 9600 consisted of initial denaturation (95°C for 5 min) followed by a step-down PCR that consisted of five cycles of denaturation (95°C, 30 s), primer annealing (61°C, 45 s), and product extension (68°C, 6 min), followed by 30 cycles of denaturation (95°C, 30 s), primer annealing (57°C, 45 s), and product extension (68°C, 6 min). This was followed by a final product extension (68°C, 10 min) for one cycle. Amplification conditions for the PTC-225 were similar except for the primer annealing temperatures during the step-down PCR, 63°C and 60°C, respectively. Amplification of KIR2DL1 and KIR3DL1 was performed using PTC-225 with the primers 2DL1E1F and 2DL1E5R and with 3DL1E3F and 3DL1E5R, respectively, as described above (Table 2). Amplification products were verified by agarose gel electrophoresis with DNA size standards. Amplification products were purified using Microcon-100 columns (Amicon, Charlotte, NC) according to the manufacturer's protocol. The purified products were resuspended in sterile water to a final concentration of 50 ng/μl.

DNA sequencing and sequence analysis

Sequencing was performed using Applied Biosystems' BigDye Terminator Ready Reaction mix according to the manufacturer's protocol (PE Applied Biosystems, Perkin-Elmer, Foster City, CA), with the minor modification of extending the extension time to 3 min. About 100–150 ng of PCR product, purified as described above, was used in each reaction. Sequencing primers for KIR2DL4 were designed to generate overlapping fragments of both strands from exon 1 through exon 9 including introns (Table 1, Fig. 1). However, due to primer annealing in exon 1, complete exon data were obtained only for exons 2–9. Sequencing primers for KIR2DL1 and KIR3DL1 are summarized in Table 2. The 5% Long Ranger gel (BMA, Rockland, ME) was electrophoresed using an ABI Prism 377 DNA sequencer (PE Applied Biosystems). Sample files were initially analyzed using Sequencher (Gene Codes, Ann Arbor, MI) software. In the families, approximately 9 kb of nucleotide sequence data for each sample was aligned with previously reported sequences (18) for identification of polymorphic sites. Segregation analysis was used to resolve heterozygous sequences and to confirm allele assignments. In the IHWS cells, sequencing of each KIR allele found in each individual was performed two to three times, with the DNA for each analysis deriving from different amplifications. Sequences were compared to a database of known KIR sequences (18) to identify alleles as homozygous or heterozygous mixtures of known alleles, or as new variants. In most IHWS cells, allele assignment was straightforward since there was a single heterozygous position. In cases of more than one heterozygous position in exons, cloning followed by

Oligonucleotide primers used for KIR2DL4 gene amplification and sequencing

	Primer	Primer sequence	Strand	Exon/intron	5′ start position[a,b]
1	2DL4E1F1	CACCCACGGTCATCATCC	Sense	E1	45
2	2DL4I6R1	CCCTTTCGCTGTTGGAGTGT	Antisense	I6	740.150
3	2DL4E6F2	CATGTTCTAGGAAACCCTTCT	Sense	E6[c]	689.2579–700
4	2DL4I9R2	TGGGCTAAGCAAAGGAGTGT	Antisense	3′–UTR	1169.127
5	2DL4E9R3	TCAGCATTTGGAAGTTCTATAC	Antisense	E9	1044
6	2DL4I1F2	CTGGAAGGGAAGGAGCAC	Sense	I1	74.8
7	2DL4E2F3	TTCTTCTTGGACCAGAGTG	Sense	E2	77
8	2DL4E2F4	TTCTTGGACCAGAGTGTGTG	Sense	E2	80
9	2DL4E2F5	TTCTTCTTGGACCAGAGTGTG	Sense	E2	77
10	2DL4I2F6	CCATCTTCACCCCAATACAA	Sense	I2	110.28
11	2DL4I2R4	CAGGCCTTCCCATGGTCAG	Antisense	I2	110.117
12	2DL4I2R5	GTCGTGGCTGTGGTTCCC	Antisense	I2	110.505
13	2DL4I2F7	GGGAACCACAGCCACGAC	Sense	I2	110.488
14	2DL4E3F8	CCAAGGTGGTCAGGACAAG	Sense	I2[c]	110.862–124
15	2DL4E3R6	GGGCCAGGCAGAGCAGAA	Antisense	E3	145
16	2DL4E3F9	CTCATTAGCCCTGTGACCC	Sense	E3	281
17	2DL4E3F10	GCAACCCCCTGGTGATCATGG	Sense	E3	369
18	2DL4E3R7	CCTCTGACCTGTGACCATG	Antisense	E3[c]	385–395.8
19	2DL4I3R8	ACAAGGAGAAGCCCAGACA	Antisense	I3	395.32
20	2DL4I3F11	TTGTAATCCTTGGAGCCTGT	Sense	I3	395.181
21	2DL4I3R9	ACAGTGGCTGGTGGAGTATC	Antisense	I3	395.390
22	2DL4I3F12	AGGGTCTGATTTCTGTCTCC	Sense	I3	395.516
23	2DL4E5F13	CCTCTTCTCCTTCCAGGTCTATATG	Sense	I3[c]	395.864–404
24	2DL4E5R10	TGGATAGATGGTAGATGT	Antisense	E5	509
25	2DL4E5F14	GGCCGACTTCCCTCTGGGT	Sense	E5	571
26	2DL4E5F15	CTTCGGCTCTTTCCATGGA	Sense	E5	619
27	2DL4I5F16	CCTCAGCCCCTCAACCTTA	Sense	I5	689.287
28	2DL4I5F17	GGCGGGAGAGTGGCTTAAA	Sense	I5	689.674
29	2DL4I5R11	CTCACTGCAACCTCCTCCT	Antisense	I5	689.717
30	2DL4I5F18	CATCTTGGCTACTGTGAACA	Sense	I5	689.1052
31	2DL4I5R12	GGGTGTAAATGCAAAGAAAAG	Antisense	I5	689.1142
32	2DL4I5R13	CCTGGCTAACATGGTGAAAC	Antisense	I5	689.1405
33	2DL4I5F19	GATTACAAGCGTGAGCCACA	Sense	I5	689.1464
34	2DL4I5R14	CAGACATTTCTCAATGAACAAA	Antisense	I5	689.1770
35	2DL4I5F20	TGGGTTGTCTCTTCTTCACTT	Sense	I5	689.1895
36	2DL4I5F21	GGGTGGATTACATCCGTGTT	Sense	I5	689.2269
37	2DL4I5R15	CATCACACTACCTGACTTAAA	Antisense	I5	689.2391
38	2DL4I6F22	GGCTTGAGAAGGGGAAGGA	Sense	I6	740.310
39	2DL4I6F23	CTGAATGTTCCAGGCAAGAA	Sense	I6	740.680
40	2DL4I6F24	GGCGGCATTCTTATCCTTTC	Sense	I6	740.1071
41	2DL4I6F25	TGTGCCCCAGGCTGGAGT	Sense	I6	740.1271
42	2DL4I6R16	GGTGGGACATGGGTGTAAT	Antisense	I6	740.1390
43	2DL4I6F26	CAGACAAGGTTTTACCATGTT	Sense	I6	740.1430
44	2DL4I6F27	GGACTTACCTCGGGGCTAA	Sense	I6	740.1867

45	2DL4I6R17	GCCTTTTCCACAGTCTCCTA	Antisense	I6	740.1975
46	2DL4I6F28	TGGGAGGCTGAGGAGGCT	Sense	I6	740.2273
47	2DL4I6R18	ACCATGTTGTCCAGGCTGAT	Antisense	I6	740.2336
48	2DL4I6F29	CTTCCTTATTTGGCTTTCTGT	Sense	I6	740.2661
49	2DL4I6F30	CCCAAGATGACAAAAGTAGCA	Sense	I6	740.3016
50	2DL4I6F31	CAGCGAGTGACAACAGAAAC	Sense	I6	740.3355
51	2DL4I6R19	CCATGTGTCCTGAGCAGTTT	Antisense	I6	740.3432
52	2DL4I6F32	AGAGGGAGAAAGGGGGTGT	Sense	I6	740.3816
53	2DL4I6F33	CTGGCAACCAAGAAATGAGA	Sense	I6	740.4163
54	2DL4E7F34	CATCTTCCTCCAGGTATCG	Sense	I6[c]	740.4231–746
55	2DL4E7R20	GCACCAGCGATGAAGGAGAA	Antisense	E7	813
56	2DL4I7R21	CGTGAGGATACAGTTCACAAT	Antisense	I7	845.205
57	2DL4I7F35	GGCAGAAAGTGGGAGATAGA	Sense	I7	845.247
58	2DL4E8R22	TGTCCCGCAGGCTCTTGGTT	Antisense	E8	879
59	2DL4E8R23	TGTTCACTGTTCTGTGTCCC	Antisense	E8	893
60	2DL4E9F36	CCTCTCTCCAGGACTCTGA	Sense	I8[c]	898.89–906
61	2DL4E9R24	ATCTGTTGAGGGTCTCTTGCT	Antisense	E9	1012
62	2DL4E9R25	CTGGGTTTGAGACAGGGCT	Antisense	E9	1132
63	2DL4E9R26	GCTGGCAAGCTGGGTTTGA	Antisense	E9	1141

E, exon; I, intron; 3′-UTR, 3′-untranslated region.
[a]Numbering of nucleotides is based on Martin et al. (3) using the genomic sequence of KIR2DL4*005 (GenBank accession number AL133414) as a reference. The codon specifying the first amino acid of the leader peptide is encoded by nucleotides 41–43, whereas the first amino acid of the mature peptide is encoded by nucleotides 104–106.
[b]Numbers refer to the 5′-position of the primer. Nucleotides were numbered so that the number to the left of the period represents the last base of the exon preceding the intron and the number to the right represents the 5′-nucleotide position in the intron. Numbering in the intron is specific to each intron. Numbers without periods indicate the position of the primer in an exon.
[c]Primer position spans both exon and intron; hence, both the start and end positions are indicated.

Table 1

sequencing was used to separate alleles. Sequences were aligned using Sequencher version 4.1.2 (Gene Codes) and/or SeqScape version 2.1.1 (Applied Biosystems). In this article, the exon numbering scheme of Garcia et al. (26) was used. Sequences were submitted to GenBank and included accession numbers: AY721618–AY721624; AY725058–AY725106; AY727757–AY727764; AY776336–AY776343. KIR allele designations for novel alleles have been assigned by the WHO Nomenclature Committee (18). The presence of interspersed repeats was analyzed using the RepeatMasker program (Smit, AFA and Green, P; RepeatMasker at http://www.repeatmasker.org).

Cloning

To confirm the presence of different lengths of the mono A repeat and to link them to other polymorphisms in eight IHWS cell lines, this region was amplified, subcloned, and sequenced. Primers 2DL4E7F34 and 2DL4I9R2 (54 and 4 in Table 1) were used to amplify a 1.2 kb fragment extending over exon 7 through exon 9 from genomic DNA with the PTC-225 thermal cycler. The cycle conditions were initial denaturation at 95°C for 5 min followed by a step-down PCR that consisted of five cycles of denaturation (95°C, 30 s), primer annealing (60°C, 45 s), and product extension (68°C, 2 min), followed by 25 cycles of denaturation (95°C, 30 s), primer annealing (57°C, 45 s), and product extension (68°C, 2 min). The purified PCR fragment was cloned into the TOPO TA cloning vector (Invitrogen Life Technologies) according to the manufacturer's instructions. Transformed (white) colonies were individually picked and transferred to a tube containing 50 µl sterile water. The contents of the tube were incubated at 99°C for 5 min to lyse the cells and to denature DNase. After a brief vortexing, 10 µl of the suspension was used for PCR amplification, as described above. Both PCR primers and four other sequencing primers (2DL4I7R21, 2DL4I7F35, 2DL4E9R3, and 2DL4E9F36; 56, 57, 5, 60 in Table 1) were used to sequence both strands.

Phylogenetic analysis

Construction of a phylogenetic tree was done using the PAUP version 4.0.0b10 (Sinaur Associates, http://www.sinauer.com, Sunderland, MA). The tree was constructed using the neighbor-joining method with a Kimura-2 parameter correction.

KIR2DL1 and KIR3DL1 oligonucleotides used for amplification and sequencing

Primer	Primer sequence (5′–3′)	Strand	Exon/intron	Start position[a,b]
2DL1 and 3DL1 PCR primers				
2DL1E1F	GGCAGCACCATGTCGCTCT	Sense	E1	32
2DL1E5R	CAAGCAGTGGGTCACTTGAC	Antisense	E5	692
3DL1E3F	CAARCCCTTCCTGTCTGCCT	Sense	E3	81
3DL1E5R	GAGAGAGAAGGTTTCTCATATG	Antisense	E5	680
2DL1 sequencing primers				
2DL1E1F	GGCAGCACCATGTCGCTCT	Sense	E1	32
2DL1i1R1	TCTAGGCCCATCACTCCATC	Antisense	I1	75.154
2DL1i1F1	GCCTGGCTACCAAGACTCAC	Sense	I1	75.872
2DL1i2R1	CCAGGGTCCCTTCTTCCTAGT	Antisense	I2	110.114
2DE4F	GGAGTCCACAGAAAACCTTC	Sense	E4[c]	110.2450–129
2DE42DR	ACCTGTGATCACGATGTCCAGA	Antisense	E4[c]	391–410.2
2DE5GF	CCTCTTCTCCTTCCAGGTCTATATG	Sense	E5[c]	410.1514–419
2DE5R	CAAGCAGTGGGTCACTTGAC	Antisense	E5	692
3DL1 sequencing primers				
3DL1E3F	CAARCCCTTCCTGTCTGCCT	Sense	I3	81
3DL1i3R	GAGGTGGGACAGTGAGAAGC	Antisense	I3	355.40
3DL1i3F	GAYGCCTTCTRAACTCACAAC	Sense	I4	355.1081
3DL1i4R	AAGTCCTRGATCATTCACTC	Antisense	I4	655.62
3DL1i4F	AAGRACCTCCCTGAGGAAAC	Sense	I4	655.1515

E, exon; I, intron; 3′-UTR, 3′-untranslated region.
[a]Numbering of nucleotides is based on Martin et al. (3) using the genomic sequence of GenBank accession numbers AY320039.1 and AC011501.8 as a reference for KIR2DL1 and KIR3DL1, respectively. The codon specifying the first amino acid of the leader peptide is encoded by nucleotides 41–43.
[b]Numbers refer to the 5′-position of the primer. Nucleotides were numbered so that the number to the left of the period represents the last base of the exon preceding the intron and the number to the right represents the 5′-nucleotide position in the intron. Numbering in the intron is specific to each intron. Numbers without periods indicate the position of the primer on an exon.
[c]Primer position spans both exon and intron; hence, both the start and end positions are indicated.

Table 2

Results

Alleles and new variants

Sequence polymorphism in the KIR2DL4 gene found in all exons and introns except exon 1 was analyzed in two families (African-American family 004 and White family 001). Eight consanguineous IHWS cell lines were also evaluated. Some of these IHWS lines (Table 3) were thought to be homozygous for KIR haplotypes based on genes present/absent (27).

Alignment of exon nucleotide sequences with known alleles of KIR2DL4 identified two alleles (2DL4*00102 and *005) and eight previously undescribed variants (2DL4*00104, *00103, *0080102, *0080103, *00802, *009, *010, and *011) (Fig. 2). Overall, six of the nine WHO-named alleles or variants of these alleles were detected. Allele KIR2DL4*005 was detected in cell lines 9042 and 9065, whereas allele KIR2DL4*00102 was detected in only one IHWS cell (9035). Based only on the coding regions, the remaining alleles were variants with one or more substitutions. One coding region sequence (labeled as 2DL4*0080103 in Fig. 2) was detected in one family cell (GU321) and three unrelated individuals (9005, 9010, and 9050). Other variants also were detected in the families as well as in the unrelated individuals. We observed three variants of the allele KIR2DL4*00202 (designated 2DL4*0080102, *0080103, and *009). GU1183 (mother of family 004) had a single allele based on the coding region of both chromosomes. But, differences in intron sequences confirmed by segregation in family 004 identified two alleles: 2DL4*0010301 was inherited by GU1181 and GU1194 and 2DL4*0010302 was inherited by GU1182 and GU1197 (Table 4). Similarly, cells GU2015 and GU321 carried an identical allele at the exon level but differed at the intron level (2DL4*0080201 in GU2015 and 2DL4*0080202 in GU321).

Synonymous and non-synonymous substitutions

Overall, there were five synonymous and three non-synonymous changes in the new variants when compared to their most similar

Distribution of 9A/10A alleles in International Histocompatibility Workshop (IHWS) cell lines

IHWS no.	IHWS name	Consanguineous[a]	Race	KIR haplotypes[b]	2DL4 alleles	Poly(A) tract
9005	Hom2	Yes	White	1/1	*0080101[c]/*0080201[c]	9A
9010	Amai	Yes	White	1/1	*0080101[c]/*0080102	9A
9023	Vavy	Yes	White	1/1	*011	9A
9030	Jhaf	Yes	White	1/1	*011	9A
9035	Jbush	Yes	White	2/2	*00102/*0010301[c]	10A[d]
9042	Tisi	Yes	White	5/5	*005	10A
9050	Mou	Yes	White	1/1	*0080101[c]	9A
9065	Hhkb	Yes	White	5/3	005/*011	9A/10A

[a]Description of consanguineous cells is according to Yang et al. (25).
[b]KIR haplotypes are based on Hsu et al. (27).
[c]Because it was difficult to clone the long amplicons, no intron data were generated for the IHWS cells. Therefore, it was not possible to determine which seven-digit allele designations should be assigned to those cells which carry the same alleles as found in the families. In these cases, the KIR Nomenclature Committee assigned the arbitrary allele designations that are shown in the table.
[d]10A was also observed for this cell line by Witt et al. (19). Unfortunately, the earlier study did not examine the other lines evaluated in this study.

Table 3

			Extracellular Ig domain												L	TM			Cytoplasmic					
Allele	Cell	Amino Acid Substitutions	Exon 3					Exon 5							Exon 6	Exon 7			Exon 9					
Codon			30	64	66	72	78	109	115	137	146	161	182	186	206	231	234	248	316	317	318	321	347	348
Nucleotide			192	293	301	317	336	429	446	514	540	584	649	659	721	796	805	845	1051	1052	1057	1066	1144	1145
2DL4*00101			A	G	A	T	A	C	A	A	C	C	A	G	T	T	T	A	A	G	G	C	T	A
00104	GU324	silent	-	-	-	-	-	-	-	-	-	-	G	-	-	C	-	-	-	-	-	-	-	-
2DL4*00102	9035		-	-	-	-	-	-	-	G	-	-	G	-	-	C	-	-	-	-	A	-	-	-
0010301 0010302	GU1183 GU1183	silent	-	-	-	-	-	-	-	G	-	-	G	-	-	C	-	-	-	-	-	-	-	-
2DL4*00201			-	-	-	-	-	-	G	-	-	-	G	C	-	-	-	-	-	-	-	-	-	-
0080201 0080202	GU2015 GU321	–1 frameshift	-	-	-	-	-	-	G	-	-	-	G	C	-	-	-	*	-	-	-	-	-	-
2DL4*00202			-	-	-	-	-	-	G	G	-	-	G	C	-	-	-	-	-	-	-	-	-	-
0080103	GU321	–1 frameshift	-	-	-	-	-	-	G	G	-	-	G	C	-	-	-	*	-	-	-	-	-	-
009	GU324	L161V –1 frameshift	-	-	-	-	-	-	G	G	-	G	G	C	-	-	-	*	-	-	-	-	-	-
0080102	9010	A317P –1 frameshift	-	-	-	-	-	-	G	G	-	-	G	C	-	-	-	*	-	C	-	-	-	-
2DL4*003			-	C	T	-	-	-	-	G	-	-	G	-	-	C	C	G	-	-	A	G	C	C
2DL4*004			-	C	-	-	-	-	G	G	-	-	G	C	-	C	-	-	-	-	-	G	C	C
2DL4*005	9042 9065		G	-	-	-	-	-	G	G	-	-	G	C	-	C	-	-	-	-	-	-	-	-
011	9023 9030 9065	silent –1 frameshift	G	-	-	-	-	-	G	G	-	-	G	C	-	-	-	*	-	-	-	-	-	-
2DL4*006			-	-	-	-	-	T	G	G	-	-	G	C	A	C	-	-	-	-	-	-	-	-
010	GU2015	P146H silent	-	-	-	-	-	T	G	G	A	-	G	C	-	C	-	-	-	-	-	-	-	-
2DL4*007			-	-	-	A	G	-	G	G	-	-	G	C	-	-	-	*	T	-	-	-	-	-

Fig. 2. **Nucleotide polymorphisms in the alleles of KIR2DL4 derived from families 001 and 004 and International Histocompatibility Workshop (IHWS) cells as compared with previously described alleles.** Only polymorphisms in the coding regions are presented. Allele KIR2DL4*00101 was used as a reference. Dashes indicate identity with the reference sequence, whereas the symbol '*' represents a deletion. The nucleotide numbering is based on Martin et al. (3). Exon and codon numbering are consistent with Garcia et al. (26) and the Immuno Polymorphism database (IPD) (http://www.ebi.ac.uk/ipd/index.html) where codon 1 is the first amino acid of the mature protein. In the IPD KIR alignment, the position of the 9A/10A deletion in allele KIR2DL4*007 differs from our positioning of the deletion (see Fig. 3). The coding sequence of KIR2DL4*0080101 is the same as that of KIR2DL4*0080103. As no intron sequence data were generated for the IHWS cells (9005, 9010, and 9050) carrying this sequence, it was not possible to determine which seven-digit allele designations should be assigned to these cells. In this case, the KIR Nomenclature Committee arbitrarily assigned the first allele in the series, KIR2DL4*0080101. In family 001, 2DL4 alleles inherited by the children include GU320 (*009/*0080202), GU322 (*009/*0080103), GU323 (*009/*0080103), GU325 (*00104/*0080103), and GU326 (*009/*0080103). In family 004, 2DL4 alleles inherited by the children include GU1181 (*0080201/*0010301), GU1182 (*0080201/*0010302), GU1194 (*010/*0010301), and GU1197 (*010/*0010302).

known alleles in addition to the frameshift discussed below (Fig. 2). Variant 2DL4*00104 had two silent substitutions (A649G and T796C), whereas 2DL4*00103 (*0010301 and *0010302), *011, and *010 contained single silent substitutions (A1057G, C796T, and A721T, respectively). All substitutions were present in other known alleles. The allele 2DL4*009 contained a non-synonymous substitution (C584G) upstream of a frameshift mutation leading to a change (L161V) in the extracellular Ig domain. Variant 2DL4*0080102 also carried a non-synonymous substitution (G1052C) at codon 317 downstream of the frameshift. Since a stop codon precedes the latter alteration, the potential amino acid change (A317P) is likely not expressed. A third apparently expressed non-synonymous substitution (C540A) in 2DL4*010 causes the alteration P146H, which is located in the region encoding the D2 domain of 2DL4*010.

Deletion in the poly(A) tract of the transmembrane coding region

Five of the eight new variants described in this study had a single nucleotide deletion in the poly(A) tract of the exon encoding a portion of the transmembrane region (Fig. 3). Five variants, 2DL4*0080102, *0080103, *00802 (*0080201 and *0080202), *009, and *011, exhibited this deletion which results in a predicted frameshift mutation. One of the four parental cell lines (GU321) and five of the eight unrelated individuals (9005, 9010, 9023, 9030, and 9050) had only $(A)_9$, whereas, two parental cell lines (GU2015 and GU324) and one unrelated individual (9065) were heterozygous $(A)_9/(A)_{10}$. Only one parental cell (GU1183) and two of the IHWS cell lines (9035 and 9042) were found to be homozygous for the allele with $(A)_{10}$.

Chromatograms covering the homopolymer region from KIR2DL4 from the parents and representative children in the two families are shown in Fig. 4. The chromatograms clearly show the segregation of the 9A and 10A alleles within the families. In cells containing alleles with differing homopolymer size in a heterozygous condition [$(A)_9$ and $(A)_{10}$] (e.g., GU324, GU325, GU2015, and GU1181), multiple overlapping peaks, marked by the arrow, occur on the 3′-end of the repeat for templates sequenced with a forward primer and on 5′-end of the repeat for templates sequenced with the reverse primer. This illustrates the impact on the alignment of the nucleotide sequences of the two alleles following a frameshift deletion in one of the alleles. The loss of alignment is not observed in homozygotes (e.g., GU321, GU320, GU1183, and GU1194).

Sequence diversity in the coding regions

Comparison of all known alleles including the new variants with KIR2DL4*00101 showed that the exon encoding, the second Ig domain (D2) was the most polymorphic with seven polymorphic positions. The cytoplasmic domain encoded by exon 9 was the second most diverse exon with six variant positions (Fig. 2). The exon encoding the extracellular domain D0 exhibited five polymorphic positions. Exons 6 and 7 exhibited one and three polymorphisms, respectively. No polymorphism was observed in exon 2 and exon 8. The number of polymorphisms in each exon tended to be proportional to the length of the exon regardless of whether the exon encoded the extracellular regions or the cytoplasmic tail of the protein. Including the two new polymorphic positions described here in exon 5 and one in exon 9, the total number of positions of sequence variability in KIR2DL4 alleles is 22.

Phylogram

To determine the relationship between named alleles and the new variants described here, a phylogram analysis based on the nucleotide sequence of the coding regions covered by exons 2–9 is depicted in Fig. 5. Roughly, four clusters are distinguishable on the phylogram tree. The largest cluster consists of alleles KIR2DL4*005, KIR2DL4*00201, KIR2DL4*00202, KIR2DL4*007 and their respective novel variants including all of the five variants containing a single nucleotide deletion in the poly(A) tract of the exon encoding the transmembrane region. KIR2DL4*007, with several nucleotide differences from other members of the subgroup, appears to be loosely clustered with this subgroup. The second cluster consists of KIR2DL4*00101, KIR2DL4*00102, and their respective variants. Features distinguishing this cluster from all others are the adenine at position 446 instead of guanine and guanine at position 659 instead of cytosine. KIR2DL4*006 and its variant branched separately, while KIR2DL4*003, the most distant allele, and KIR2DL4*004, the second most distant allele, clustered together. The distinguishing feature of the cluster containing KIR2DL4*006 and 2DL4*010 is a thymine at position 429 instead of cytosine. The group of KIR2DL4*003 and KIR2DL4*004 is characterized by the changes G293C (C = other alleles/G = KIR2DL4*003/*004), C1066G, T1144C, and A1145C.

Lack of KIR homozygosity in consanguineous individuals

Eight consanguineous IHWS cells, known to be homozygous for HLA loci on chromosome 6, were studied. All of these cell lines, except 9065, had been thought to be homozygous for KIR haplotypes based on genes present/absent in an earlier study by Hsu et al. (Table 3) (27), although subsequent studies observed heterozygosity for KIR3DL1/3DS1 and 3DL2 in cell lines 9030, 9035, and 9050 (28, 29). While cell lines 9023, 9030, 9042, and 9050 appeared to carry single

Nucleotide polymorphisms in the KIR2DL4 introns[a]

Alleles (cells)	Intron position[b]	005 – AL133414	00104 (GU324)	0010301 (GU1183)	0010302 (GU1183)	0080202 (GU321)	0080201 (GU2015)	0080103 (GU321)	009 (GU324)	010 (GU2015)
Intron 1	74.161	**C**	G	G	G	–	G	G	G	–
	74.180	**C**	A	A	A	A	A	A	A	A
Intron 2	110.59	**G**	–	–	–	A	–	–	–	–
	110.186	**C**	–	–	–	–	T	T	T	T
	110.214–47	**$(34N)_3$**	$(34N)_3$	$(34N)_4$	$(34N)_4$	$(34N)_3$	$(34N)_3$	$(34N)_3$	$(34N)_3$	$(34N)_3$
	110.269	**T**	–	C	C	–	–	C	C	–
	110.437	**G**	–	C	–	–	–	–	–	–
	110.650	**G**	–	–	–	C	–	–	–	–
	110.745	**C**	T	–	–	–	–	–	–	–
Intron 3	395.171	**G**	–	A	–	–	–	–	–	–
	395.288	**C**	T	T	T	T	T	T	T	T
	395.568	**C**	T	T	T	–	–	–	–	–
	395.601	**T**	A	A	A	A	A	A	A	A
	395.744	**G**	–	–	–	–	A	A	A	–
	395.774–779	**$(AG)_{11}$**	$(AG)_8$	$(AG)_8$	$(AG)_8$	$(AG)_8$	$(AG)_8$	$(AG)_8$	$(AG)_8$	$(AG)_8$
Intron 5	689.123	**T**	C	C	C	–	–	–	–	–
	689.486	**G**	–	–	C	–	–	–	–	–
	689.501	**C**	–	–	–	T	–	–	–	–
	689.526	**C**	–	–	–	–	T	T	T	–
	689.572	**C**	T	T	T	T	T	T	T	T
	689.644	**G**	C	–	–	C	–	C	C	C
	689.743	**C**	–	–	–	–	–	–	–	T
	689.771–4	**$(4N)_4$**[c]	$(4N)_4$	$(4N)_5$	$(4N)_5$	$(4N)_4$	$(4N)_4$	$(4N)_4$	$(4N)_4$	$(4N)_4$
	689.930	**C**	–	T	T	–	–	–	–	–
	689.946	**C**	–	T	–	–	–	–	–	–
	689.961	**T**	A	A	A	A	A	A	A	A
	689.972	**G**	–	C	C	–	–	–	–	–
	689.1009	**A**	–	–	–	T	–	–	–	–
	689.1354	**C**	A	A	A	–	–	–	–	–
	689.1465	**T**	C	–	–	C	C	–	–	–
	689.1483	**C**	–	–	–	A	–	–	–	–
	689.1504	**T**	C	C	C	–	–	–	–	–
	689.1578	**C**	–	–	–	T	–	–	–	–

Table 4 **continued overleaf**

Continued

Alleles (cells)	Intron position[b]	005 – AL133414	00104 (GU324)	0010301 (GU1183)	0010302 (GU1183)	0080202 (GU321)	0080201 (GU2015)	0080103 (GU321)	009 (GU324)	010 (GU2015)
	689.1905	**G**	**–**	T	T	–	–	–	–	–
	689.2305	**C**	**–**	T	T	–	–	–	–	–
	689.2505	**T**	**–**	G	G	–	–	–	–	–
	689.2547	*	*	AA	AA	*	*	*	*	*
	689.2574	**C**	**–**	A	–	–	–	–	–	–
	689.2586	**A**	**–**	–	–	G	–	–	–	–
Intron 6	740.143	**C**	**–**	–	–	–	–	–	–	G
	740.329	C	A	A	A	A	A	A	A	A
	740.354	A	G	G	G	–	–	–	–	–
	740.531	C	T	T	T	–	–	–	–	–
	740.589	A	T	T	T	T	T	T	T	T
	740.1058	C	–	–	–	T	T	T	T	–
	740.1175	G	–	A	A	–	–	–	–	–
	740.1244	T	–	–	–	–	–	*	*	–
	740.1259	G	–	T	T	A	T	–	–	T
	740.1268	C	T	–	–	T	T	–	–	–
	740.1519	C	A	A	A	A	A	A	A	A
	740.1748	G	–	–	–	–	C	C	C	–
	740.1764	G	C	C	C	C	–	–	–	C
	740.1765	A	G	–	–	G	G	–	–	–
	740.1790	G	–	–	–	–	–	–	–	A
	740.1832	G	–	–	–	C	–	–	–	C
	740.1842	C	–	–	–	–	–	–	–	T
	740.1853	C	T	T	T	T	–	–	–	–
	740.1864	A	C	–	–	C	C	–	–	–
	740.2086	C	G	G	G	G	G	G	G	G
	740.2387	A	G	G	G	G	–	–	–	–
	740.2409	G	–	A	A	–	–	–	–	–
	740.2482	C	–	–	T	–	–	–	–	–
	740.2638	A	–	–	–	G	G	G	G	G
	740.2720	G	A	A	A	A	A	A	A	A
	740.2756	C	–	–	–	–	G	G	G	–
	740.3263	G	C	C	C	–	–	–	–	C

	740.3299	G	–	T	T	–	–	–	–	–
	740.3693	T	G	G	G	–	–	–	–	G
	740.3902	C	–	A	A	–	–	–	–	A
	740.4138	C	–	–	–	T	T	T	T	–
	740.4230	T	–	–	–	C	C	C	C	–
Intron 7	845.149	A	C	C	C	C	C	C	C	C
	845.177	A	–	–	–	G	G	G	G	–
	845.440	C	T	T	T	T	T	T	T	T
Intron 8	898.43	C	–	–	–	G	G	G	G	–
	898.44	T	–	–	–	A	A	A	A	–

[a]Allele KIR2DL4*005 (GenBank accession number AL133414) was used as a reference. Dashes indicate identity with the reference sequence, whereas the symbol '*' represents deletions.

[b]The nucleotide numbering is based on Martin et al. (3). The exons/introns are numbered based on an alignment of all of the KIR gene sequences as reported in Garcia et al. (26). Although exon 4 is missing in 2DL4, we have labeled the transmembrane exon as exon 7 in order to keep the relationship among KIR exons in the different genes consistent. Nucleotide positions were numbered in such a way that the number to the left of the period represents the last base of the exon preceding the intron and the number to the right represents the nucleotide position in the intron. Numbering in the intron is specific to each intron. For instance, 74.111 refers to 74 as the last nucleotide position of exon 1, while nucleotide number 111 is in intron one. Similarly, 110.59 refers to 110 as the last position of exon 2 and 59 as the 59th position of intron two.

[c]4N represents (AAAT).

Table 4

	Exon7 (partial) ▼					Intron 7 (partial)					▼ Exon 8 (partial)				
Consensus	TCC	AAA	AAA	AAA	A	gtaag	cctca	..//..	tctcc	tacag	AT	GCT	GCT	GTA	ATG
KIR2DL4*00101	---	---	---	---	-						--	---	---	---	---
KIR2DL4*00102	---	---	---	---	-						--	---	---	---	---
KIR2DL4*0010301	---	---	---	---	-	-----	-----	..//..	-----	-----	--	---	---	---	---
KIR2DL4*0010302	---	---	---	---	-	-----	-----	..//..	-----	-----	--	---	---	---	---
KIR2DL4*00104	---	---	---	---	-	-----	-----	..//..	-----	-----	--	---	---	---	---
KIR2DL4*00201	---	---	---	---	-						--	---	---	---	---
KIR2DL4*00202	---	---	---	---	-						--	---	---	---	---
KIR2DL4*003	---	---	---	---	G						--	---	---	---	---
KIR2DL4*004	---	---	---	---	-						--	---	---	---	---
KIR2DL4*005	---	---	---	---	-						--	---	---	---	---
KIR2DL4*006	---	---	---	---	-						--	---	---	---	---
KIR2DL4*007	---	---	---	---	*						--	---	---	---	---
KIR2DL4*0080101	---	---	---	---	*						--	---	---	---	---
KIR2DL4*0080102	---	---	---	---	*						--	---	---	---	---
KIR2DL4*0080103	---	---	---	---	*	-----	-----	..//..	-----	-----	--	---	---	---	---
KIR2DL4*0080201	---	---	---	---	*	-----	-----	..//..	-----	-----	--	---	---	---	---
KIR2DL4*0080202	---	---	---	---	*	-----	-----	..//..	-----	-----	--	---	---	---	---
KIR2DL4*009	---	---	---	---	*	-----	-----	..//..	-----	-----	--	---	---	---	---
KIR2DL4*010	---	---	---	---	-	-----	-----	..//..	-----	-----	--	---	---	---	---
KIR2DL4*011	---	---	---	---	*						--	---	---	---	---

Fig. 3. **Alignment of nucleotide sequence of KIR2DL4 alleles in the region of exon 7/intron 7/exon 8.** The nucleotide numbering is based on Martin et al. (3). Exon and codon numbering is consistent with Garcia et al. (26) and the Immuno Polymorphism database (http://www.ebi.ac.uk/ipd/index.html) where codon 1 is the first amino acid of the mature protein. Although exon 4 is missing in 2DL4, we have labeled the transmembrane exon as exon 7 in order to keep the relationship among killer-cell immunoglobulin-like receptor (KIR) exons from different genes consistent. Boldface indicates novel alleles detected in this study. Dashes indicate identity with the reference sequence, whereas the symbol '*' represents a deletion. Upper case letters, exon sequence; lower case letters, intron sequence; blank rows represent sequence data not available; a triangle indicates the end of exon 7 and the beginning of exon 8.

alleles of KIR2DL4 based on their sequences, cell lines 9005, 9010, 9035, and 9065 were found to each carry two almost identical alleles, differing by a single nucleotide and/or a deletion in the A homopolymer (Table 3). As some chromosome pairs in consanguineous individuals are identical by descent, the presence of almost identical alleles in the heterozygotes prompted the analysis of adjacent KIR loci. KIR3DL1 and KIR2DL1 flank 2DL4 on chromosome 19 (1). The coding regions of KIR3DL1 including exons 3–5 and KIR2DL1 including exons 1–5 were sequenced from cell lines 9005, 9010, and 9035. [Cell line 9065 was suggested to carry two non-identical chromosomes

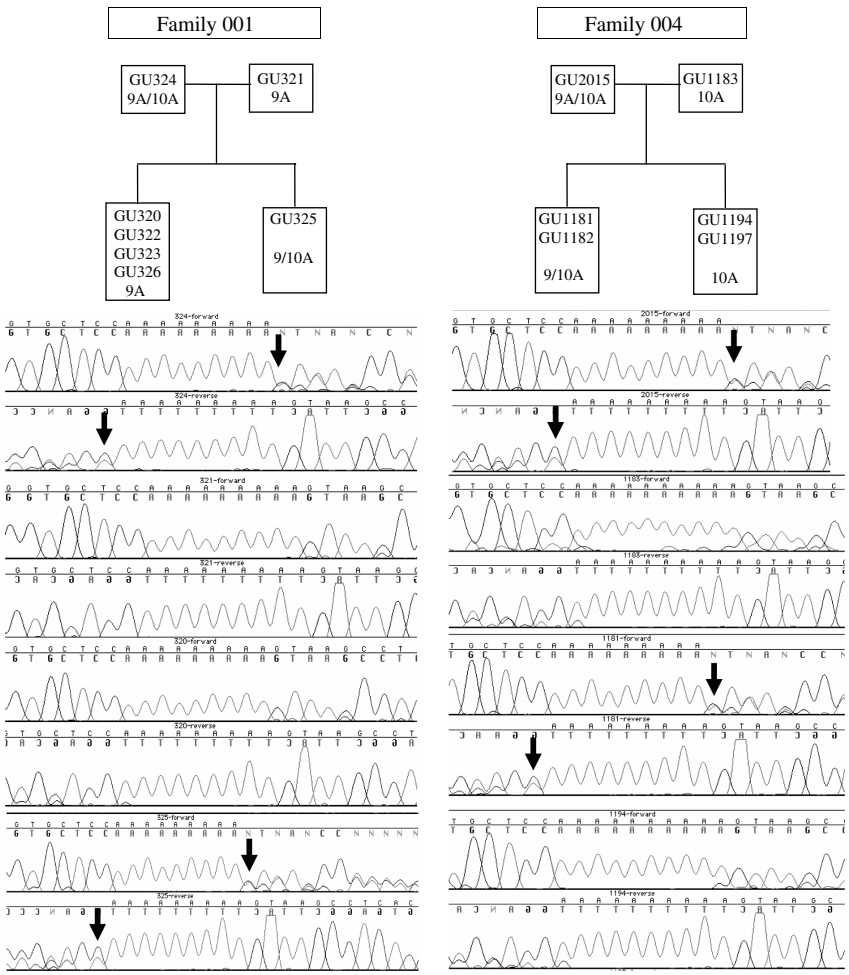

Fig. 4. **Family cells containing 9A and 10A variations of KIR2DL4 and representative sequencing chromatograms.** Family trees identify the parents and children of the families used in this study. Siblings with the same KIR2DL4 homopolymer variations are shown in the same box. Sequencing chromatograms covering both DNA strands in the region of the homopolymer segment of the sequence for representative family members are shown. The chromatograms show the exon 7–intron 7 boundary. Arrows indicate the start of multiple overlapping peaks which are produced by differences in the length of the A homopolymer leading to a difference in the reading frame of the two alleles.

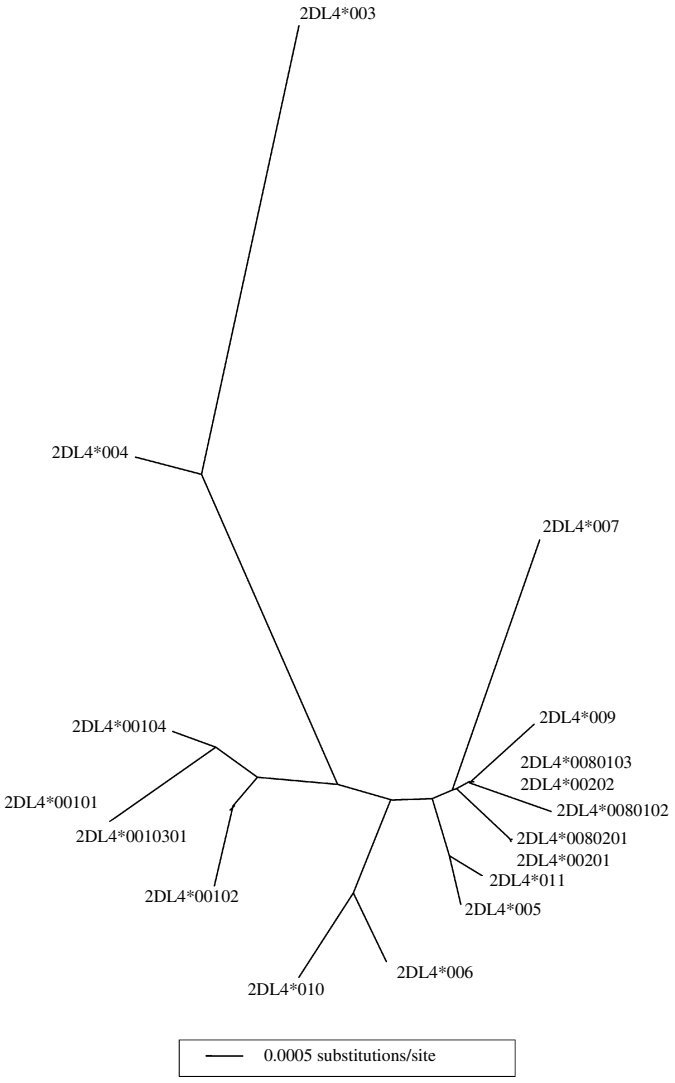

Fig. 5. **Phylogram based on nucleotide sequences of the coding regions of exons 2–9 of known alleles and new variants of KIR2DL4.** A neighbor-joining tree with Kimura-2 correction was used to generate distances. Phylogram branch lengths represent actual distances between sequences.

based on analysis of KIR gene content (27).] Sequence analysis of KIR3DL1 revealed several heterozygous positions in exon 3 and exon 4 defining two alleles in the cell lines 9005 (KIR3DL1*00101/KIR3DL1*00401 or KIR3DL1*00101/KIR3DL1*00402) and 9035 (KIR3DL1*002 or KIR3DL1*003 or KIR3DL1*007 and KIR3DL1-*008). [Probe based testing has assigned KIR3DL1*002/KIR3DL1*008 to cell 9035 (28, 29).] However, no heterozygous positions were detected in cell line 9010 (KIR3DL1*00101). Further sequencing of locus KIR2DL1 showed that cell line 9010 is heterozygous for this locus carrying alleles KIR2DL1*001 and a novel 001 variant. The 001v of 2DL1 contains a non-synonymous substitution (C150G) that results in an amino acid change from proline to arginine at codon 16. Our result is in agreement with the finding of Keaney et al. (30) who reported two KIR2DL3 alleles, KIR2DL3*002 and KIR2DL3*005, in cell line 9010 suggesting that this cell carries non-identical chromosomes.

Sequence diversity in the non-coding region

In order to investigate the nature of sequence variation in the non-coding regions and because of the scarcity of intron data available in GenBank, complete intron sequence data covering introns 1–8 were generated for both families. The complete KIR2DL4 genomic sequence of KIR2DL4*005, denoted by GenBank accession number AL133414, was used as a reference. Except for an insertion of three dinucleotide repeats (AG) at position 779 of intron 3 (Table 4), AL133414 has the same intron lengths as the majority of the genomic sequences isolated from the families. Two alleles, 2DL4*0080103 in cell GU321 and 2DL4*009 in cell GU324, had a single thymine deletion at position 740.1244 in intron 6. The two alleles, 2DL4*0010301 and 2DL4*0010302, detected in cell GU1183 had three insertions, including an insertion of a 34-nucleotide repeat in intron 2 (position 110.214) and insertions of a tetranucleotide (position 689.771) and a mononucleotide (position 689.2547) repeat both in intron 5. While both chromosomes in cell GU1183 contained four tandem repeats, all other investigated individuals have three $(N)_{34}$ repeats – (GAGTCTCTCATGAACTAGTAAGAGGAGATC(C/T)TGG) – in intron 2.

Although six different coding region sequences were defined in the families, eight alleles were found based on both exon and intron sequences. It is worth noting that the two alleles from individual GU1183 (2DL4*0010301 and 2DL4*0010302) were identical at the exon level but exhibited six differences at the intron level – a single difference each in introns 2, 3, and 6 and three differences in intron 5. Similarly, cells GU2015 and GU321 had the identical sequence at the exon level. However, there were 19 nucleotide differences in the introns (one each in introns 1 and 3, three in intron 2, and seven each in introns 5 and 6) distinguishing 2DL4*0080201 and 2DL4*0080202 (Table 4). The remaining parental alleles were diverse both in the exons and introns. The larger introns (introns 5 and 6) contain the highest number of polymorphic positions, 24 and 32 positions, respectively, together accounting for 74% of the total intron polymorphisms.

Repetitive sequences

In order to assess the nature and distribution of interspersed repeats and low complexity DNA sequences in the KIR2DL4 gene, one of the alleles (2DL4*0080201) derived from cell GU2015 was analyzed using the RepeatMasker program. Overall, interspersed repeats comprised

46% of the KIR2DL4 gene sequence. The analysis showed the presence of six SINEs (13.73% of the total length), one LINE (12.41%), and five LTR elements (19.51%). The distribution of the 12 elements in the introns was as follows: intron 3 contained two LTR elements (ERVL and MaLR), intron 5 contained two SINEs (both of which were Alu) and one LINE (L1), and intron 6 contained four SINEs and three LTRs (Table 5).

Discussion

We investigated the level of KIR2DL4 nucleotide variation in two families and a panel of unrelated individuals by DNA sequencing. Based on coding region segments, two known alleles and eight new variants were identified. The new variants differed from the known alleles by one (*00103, *0080103, and *00802) or two (*00104, *0080102, *009, *010, and *011) nucleotides. Five of the eight new variants contained a single nucleotide deletion in a region containing a poly(A) tract that results in a frameshift mutation and likely produces a truncated protein without a cytoplasmic tail.

Allelic polymorphism

In addition to haplotype diversity defined by gene number and content, KIR diversity is also expanded by allelic polymorphism. Considering the combination of these two levels of diversity, the likelihood that two unrelated individuals have identical KIR haplotypes is very low. Recent sequence analysis of the KIR region on chromosome 19 (14) strengthened the postulation that inter- and intralocus recombination together with other events including point mutations, gene duplications, and deletions may have brought about the rapid evolution of KIR gene diversity (7, 14). In our study, we observed several putative point mutations as well as deletions in KIR2DL4 genomic DNA derived from related and unrelated individuals. However, there is a possibility that some of the variants may have derived from intralocus recombination as most of the variation is shared by other 2DL4 alleles. Our allele assignments in the IHWS cells are consistent with those made by a previous study using sequence-specific oligonucleotide probe hybridization (16). However, since the previous study was designed to explore existing polymorphisms only in the region encoding the extracellular domains, it could not detect the new polymorphisms reported in this study. Thus, we identified variation of allele KIR2DL4*00202 in 9010, KIR2DL4*005 in 9030, KIR2DL4*00102 in 9035, and KIR2DL4*00202 in 9050, differences missed by the SSOP report (16).

Deletion in the poly(A) tract of the transmembrane region

The exon encoding the transmembrane region (exon 7, see Fig. 3 for an explanation of exon numbering) ends with a mononucleotide repeat $(A)_{10}$, in all but two alleles assigned in the KIR nomenclature report (18). In KIR2DL4*003, the last nucleotide of the exon, the tenth A, has been replaced by a G encoding an aspartic acid (forming a GAT codon with the AT contributed by exon 8) instead of an

Distribution of interspersed repeat elements in KIR2DL4 gene

	Element	Family	Class	Length (Nt)	Location (intron)	% of intron
1	LTR33A	ERVL	LTR	139	3	15.8
2	MLT1D	MaLR	LTR	474	3	53.9
3	AluSx	Alu	SINE	303	5	11.7
4	LIMA4	L1	LINE	1306	5	50.3
5	AluSq	Alu	SINE	304	5	11.7
6	MIRb	MIR	SINE	110	6	2.6
7	MLT1D	MaLR	LTR	562	6	13.2
8	AluSx	Alu	SINE	301	6	7.1
9	AluSx	Alu	SINE	291	6	6.8
10	MIRb	MIR	SINE	136	6	3.2
11	MER70B	ERVL	LTR	439	6	10.3
12	MSTB1	MaLR	LTR	439	6	10.3
Total				4804		50.7[a]

Nt, nucleotide.
[a]Total intron length was 9465 nucleotides based on allele 2DL4*0080201 from GU2015; total gene length was 10,525 nucleotides.

Table 5

asparagine (AAT) at this position. KIR2DL4*007, in contrast, has a deletion of one of the adenines in the repeat with no subsequent nucleotide replacement resulting in a predicted frameshift alteration that introduces a premature stop codon just four residues into the cytoplasmic tail resulting in a truncated protein that lacks the ITIM motif. In addition to KIR2DL4*007, other allelic variants of 2DL4 without a WHO nomenclature designation containing this frameshift mutation were observed by Witt et al. (19). Further investigation of the functional role of the deletion variant revealed that the protein encoded by the 9A variant showed undetectable or reduced membrane expression in peripheral blood NK cells (10, 21). In this study, we report that five of the eight new variants had a single nucleotide deletion in the poly(A) tract of the exon encoding the transmembrane region. Witt et al. detected the 9A allelic variant in 14 of 23 IHWS cell lines investigated. They had suggested that a single allele of 2DL4 carried the 9A sequence and that this allele was found at a gene frequency of 0.53 (19), although a later study by the same investigators reported multiple alleles with the single adenine deletion (20). Our data suggest that the 9A variation is common among many 2DL4 alleles so the high frequency is likely to be the result of the 9A variation appearing throughout the 2DL4 allele family. The high frequency of a variation which may result in loss of an inhibitory receptor from the NK cell surface may be the result of selection for a reduced number of inhibitory receptors to lower the activation threshold for recognition of pathogens (31).

The deletion of a single adenine in the five variants was confirmed by direct sequencing as well as by subcloning and sequencing of at least two PCR products derived from the region flanking the poly(A) tract in the IHWS cell lines. Both strands of a 1.2 kb fragment extending over exon 7 through exon 9 were sequenced in two families of 13 individuals and in eight unrelated individuals. One of the parents in each of the two families studied contained the allele with a single adenine deletion. Mendelian segregation of the truncated alleles in both families as well as consistent results of multiple amplifications and cloning in the unrelated individuals strongly suggest that the deletion is real and did not result from *in vitro* amplification artifacts. This study does not exclude, however, the possibility of cell culture artifacts that may have arisen from multiple passages of the cell lines over the years. Changes in the length of the homopolymer sequence might have occurred during evolution by errors in DNA replication. In their extensive review on DNA replication fidelity, Kunkel and Bebenek (32) described slippage of the DNA polymerase during replication of homopolymeric sequences such as poly(A) repeats leading to insertion or deletion of a single nucleotide. In other genes, DNA replication errors, particularly in coding region repeat sequences, have been noted as instabilities in microsatellite repetitive sequences (33, 34). Deletion variants are not uncommon in other KIR genes such as KIR2DS4 (22) and are thought to be one of the mechanisms that produce allelic diversity.

It should be noted, however, that a few clones of KIR2DL4 alleles obtained during analysis showed variation in the length of the homopolymer repeat not found in the original heterozygous sequence, suggesting that this region can be subject to artifacts that arise during PCR amplification and cloning. For example, in cell 9065, clones carried from 7A to 10A in the homopolymer stretch. In a measurement of Taq DNA polymerase slippage mutation rates in templates containing dinucleotide and mononucleotide repeats, Shinde et al. (35) observed higher mutation rates for $(A)_n$ compared to $(CA)_n$ tracts and mutation was observed for $(A)_n$ tracts with more than eight repeats. Thus, care should be taken in the assignment of the length of the homopolymer A stretch to avoid sequence artifacts.

The effect of non-synonymous substitutions

The occurrence of two non-synonymous changes in an extracellular Ig-like domain suggests that these changes might play a role in altering the interaction of KIR2DL4 with its ligand. Analysis of the KIR/HLA interface in KIR2DL1-3 revealed that three of the six loops found in the ligand-binding domain of KIR interact with the α2 helix of HLA (36). The KIR/HLA interface is characterized by charge complementarity in which KIR and HLA contribute six acidic (four in the D2 domain) and six basic residues, respectively, which result in the formation of salt bridges. Even though amino acid sequence alignment showed that the D2 domain of KIR2DL4 is 78–82% identical to that of KIR2DL1-3, the roles of substitutions L161V and P146H are difficult to predict from the published structures of KIR/HLA complexes. Sequence and structure alignment predicts that P146H and L161V of KIR2DL4 correspond to Pro/Ser151 and Leu166 of KIR2DL1-3, respectively (unpublished observations). Both Pro/Ser151 and Leu166 are not located in the loops found in the ligand-binding domain of KIR (36). Furthermore, since the putative ligand of KIR2DL4 might be different from the classical HLA molecules, the residues involved in the binding and, consequently, the characteristics of the interface could be different. Homology modeling of the 3D structure of KIR2DL4 and docking of the potential ligand, once found, may provide an insight into the role of allelic diversity in the binding of KIR2DL4 to its ligand.

Heterozygosity in IHWS cells

The presence of two almost identical alleles in cells potentially carrying two copies of chromosome 19 that were identical by descent was intriguing. Evaluation of surrounding loci (KIR2DL1 and

KIR3DL1) suggested that all three investigated IHWS cell lines (9005, 9010, and 9035) were heterozygous for nearby loci and potentially carried non-identical copies of the KIR-bearing chromosome. Whether these closely related KIR2DL4 alleles arose from a recent genetic event is not known, but the observation is consistent with earlier studies on the characteristics of the genomic DNA which suggested rapid evolution of the KIR complex (7, 14).

Intron sequences

This study also surveyed the nature of nucleotide variation in the non-coding regions of KIR2DL4. While the polymorphisms in the introns are unlikely to alter the gene function, intron sequences can be utilized to design PCR amplification reagents for KIR typing so that the full sequence of the coding region can be accessible for assays. Although the extent of interlocus intron sequence conservation for KIR2DL4 and other KIR loci is unknown, alignment of all genomic sequences provided by this study showed extended conserved regions in the introns which might be utilized to design a set of KIR2DL4-specific primers to amplify exons of interest.

Analysis of the nature and location of ancestral and postduplication retroelements in the KIR genes provides insight into the evolution of the KIR gene family. Retrotransposons (LTR and non-LTR) are one of the two major classes of mobile elements. In this study, RepeatMasker analysis showed that the type, position, and number of Alus and other elements are consistent with the previous report in which Alu, LINEs, and LTRs accounted for 30% of the KIR region in the PAC 1060P11 clone (7, 15). The abundance of AluS (new) and the absence of AluJ (old) sequences in KIR genes support the theory of recent origin of the KIR gene family. The fact that AluSs are similarly located in all KIR genes suggested that the KIR loci were derived by duplication of a single primordial locus (15). Based on a thorough examination of retroelements in the KIR gene family, Martin et al. (3) proposed that 2DL4 is one of the oldest KIR genes.

In summary, KIR2DL4 is unique in many ways from other KIR molecules possessing characteristics of both activation and inhibitory receptors, carrying a unique extracellular domain (D0), and being expressed by all, not a subset of, NK cells. Like other KIRs, the KIR2DL4 locus encodes multiple allelic variants that differ subtly from one another. One interesting feature of the allelic diversity is the variation in the length of a homopolymer region resulting in potentially truncated molecules associated with many allelic types. The functional impact of this diversity remains to be defined but it may impact the activation threshold required to recognize pathogens by NK cells.

References

1. Carrington M, Norman P. The KIR gene cluster. Bethesda, MD: National Library of Medicine (US), NCBI, 2003.
2. Vilches C, Parham P. KIR: diverse, rapidly evolving receptors of innate and adaptive immunity. *Annu Rev Immunol* 2002: **20**: 217–51.
3. Martin AM, Freitas EM, Witt CS, Christiansen FT. The genomic organization and evolution of the natural killer immunoglobulin-like receptor (KIR) gene cluster. *Immunogenetics* 2000: **51**: 268–80.
4. Selvakumar A, Steffens U, Dupont B. NK cell receptor gene of the KIR family with two IG domains but highest homology to KIR receptors with three IG domains. *Tissue Antigens* 1996: **48**: 285–94.
5. Vilches C, Rajalingam R, Uhrberg M, Gardiner CM, Young NT, Parham P. KIR2DL5, a novel killer-cell receptor with a D0–D2 configuration of Ig-like domains. *J Immunol* 2000: **164**: 5797–804.
6. Valiante NM, Uhrberg M, Shilling HG et al. Functionally and structurally distinct NK cell receptor repertoires in the peripheral blood of two human donors. *Immunity* 1997: **7**: 739–51.
7. Trowsdale J, Barten R, Haude A, Stewart CA, Beck S, Wilson MJ. The genomic context of natural killer receptor extended gene families. *Immunol Rev* 2001: **181**: 20–38.
8. Faure M, Long EO. KIR2DL4 (CD158d), an NK cell-activating receptor with inhibitory potential. *J Immunol* 2002: **168**: 6208–14.
9. Rajagopalan S, Fu J, Long EO. Cutting edge: induction of IFN-gamma production but not cytotoxicity by the killer cell Ig-like receptor KIR2DL4 (CD158d) in resting NK cells. *J Immunol* 2001: **167**: 1877–81.
10. Kikuchi-Maki A, Yusa S, Catina TL, Campbell KS. KIR2DL4 is an IL-2-regulated NK cell receptor that exhibits limited expression in humans but triggers strong IFN-gamma production. *J Immunol* 2003: **171**: 3415–25.
11. Yusa S, Catina TL, Campbell KS. SHP-1- and phosphotyrosine-independent inhibitory signaling by a killer cell Ig-like receptor cytoplasmic domain in human NK cells. *J Immunol* 2002: **168**: 5047–57.
12. Rajagopalan S, Long EO. A human histocompatibility leukocyte antigen (HLA)-G-specific receptor expressed on all natural killer cells. *J Exp Med* 1999: **189**: 1093–9.
13. Boyson JE, Erskine R, Whitman MC et al. Disulfide bond-mediated dimerization of HLA-G on the cell surface. *Proc Natl Acad Sci USA* 2002: **99**: 16180–5.
14. Martin A, Kulski J, Gaudieri S et al. Comparative genomic analysis, diversity and evolution of two KIR haplotypes A and B. *Gene* 2004: **335**: 121–31.
15. Wilson MJ, Torkar M, Haude A et al. Plasticity in the organization and sequences of human KIR/ILT gene families. *Proc Natl Acad Sci USA* 2000: **97**: 4778–83.
16. Williams F, Meenagh A, Sleator C, Middleton D. Investigation of killer cell immunoglobulin like receptor gene diversity: I. KIR2DL4. *Hum Immunol* 2004: **65**: 31–8.
17. Martin MP, Bashirova A, Traherne J, Trowsdale J, Carrington M. Cutting edge: Expansion of the KIR locus by unequal crossing over. *J Immunol* 2003: **171**: 2192–5.

18. Marsh SG, Parham P, Dupont B et al. Killer-cell immunoglobulin-like receptor (KIR) nomenclature report, 2002. *Hum Immunol* 2003: **64**: 648–54.
19. Witt CS, Martin A, Christiansen FT. Detection of KIR2DL*4* alleles by sequencing and SSCP reveals a common allele with a shortened cytoplasmic tail. *Tissue Antigens* 2000: **56**: 248–57.
20. Witt CS, Whiteway JM, Warren HS et al. Alleles of the KIR2DL*4* receptor and their lack of association with pre-eclampsia. *Eur J Immunol* 2002: **32**: 18–29.
21. Goodridge JP, Witt CS, Christiansen FT, Warren HS. KIR2DL*4* (CD158d): genotype influences expression and function in NK cells. *J Immunol* 2003: **171**: 1768–74.
22. Maxwell LD, Wallace A, Middleton D, Curran MD. A common KIR2DS4 deletion variant in the human that predicts a soluble KIR molecule analogous to the KIR1D molecule observed in the rhesus monkey. *Tissue Antigens* 2002: **60**: 254–8.
23. Pando MJ, Gardiner CM, Gleimer M, McQueen KL, Parham P. The protein made from a common allele of KIR3DL1 (3DL1*004) is poorly expressed at cell surfaces due to substitution at positions 86 in Ig domain 0 and 182 in Ig domain 1. *J Immunol* 2003: **171**: 6640–9.
24. Vilches C, Gardiner CM, Parham P. Gene structure and promoter variation of expressed and nonexpressed variants of the KIR2DL5 gene. *J Immunol* 2000: **165**: 6416–21.
25. Yang SY, Milford E, Hammerling U, Dupont B. Description of the reference panel of B-lymphoblastoid cell lines for factors of the HLA system: the B-cell line panel designed for the Tenth International Histocompatibility Workshop. In: Dupont B, ed. Immunology of HLA, **Vol. I**. Histocompatibility testing 1987. New York: Springer-Verlag, 1989: 11–9.
26. Garcia CA, Robinson J, Guethlein LA, Parham P, Madrigal JA, Marsh SG. Human KIR sequences 2003. *Immunogenetics* 2003: **55**: 227–39.
27. Hsu KC, Liu XR, Selvakumar A, Mickelson E, O'Reilly RJ, Dupont B. Killer Ig-like receptor haplotype analysis by gene content: evidence for genomic diversity with a minimum of six basic framework haplotypes, each with multiple subsets. *J Immunol* 2002: **169**: 5118–29.
28. Halfpenny IA, Middleton D, Barnett YA, Williams F. Investigation of killer cell immunoglobulin-like receptor gene diversity: IV. KIR3DL1/S1. *Hum Immunol* 2004: **65**: 602–12.
29. Meenagh A, Williams F, Sleator C, Halfpenny IA, Middleton D. Investigation of killer cell immunoglobulin-like receptor gene diversity V. KIR3DL2. *Tissue Antigens* 2004: **64**: 226–34.
30. Keaney L, Williams F, Meenagh A, Sleator C, Middleton D. Investigation of killer cell immunoglobulin-like receptor gene diversity III. KIR2DL3. *Tissue Antigens* 2004: **64**: 188–94.
31. Khakoo SI, Thio CL, Martin MP et al. HLA and NK cell inhibitory receptor genes in resolving hepatitis C virus infection. *Science* 2004: **305**: 872–4.
32. Kunkel TA, Bebenek R. DNA replication fidelity. *Annu Rev Biochem* 2000: **69**: 497–529.
33. Duval A, Hamelin R. Mutations at coding repeat sequences in mismatch repair-deficient human cancers: Toward a new concept of target genes for instability. *Cancer Res* 2002: **62**: 2447–54.
34. Li YC, Korol AB, Fahima T, Beiles A, Nevo E. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol Ecol* 2002: **11**: 2453–65.
35. Shinde D, Lai YL, Sun FZ, Arnheim N. Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)(n) and (A/T)(n) microsatellites. *Nucleic Acids Res* 2003: **31**: 974–80.
36. Boyington JC, Sun PD. A structural perspective on MHC class I recognition by killer cell immunoglobulin-like receptors. *Mol Immunol* 2002: **38**: 1007–21.